

New Speaker:

- ATRAC 2.4 = plain text entry into metadata, this will liberate the user from using xml

New Speaker:

- Most of the presentations posted under the EDM wiki (NCDC wiki not externally accessible)
- NOAA data management directives
 - Data will be: discoverable
 - Accessible
 - Documented
 - Preserved- this is an interesting issue- seems like there is some trashing
 - For all types of users and applications
- NOAA environmental data management
 - Principles
 - Full and open access
 - Data preservation
 - Information quality (metadata)
 - Ease of use (formats)
 - Governance: many groups
 - EDMC procedural Directives: apply NOAA wide
 - data management planning- prior to data collection: template
 - Archive procedure: what and how to submit
 - Data documentation: ISO
 - Data sharing by NOAA grantees: state in proposal how to share and must share within 2 years- legal agreement
 - New ones: data citation: unique identifier- maybe DOI
 - Data access: what online services to use so data can be obtained
 - Resources
 - Budget
 - Project specific
 - NOAA wide
 - personnel
 - Standards
 - Interoperability, etc
 - Architecture
 - Service-based, system of systems approach: envisioning portals, ancillary services
 - Archival data centers
 - CLASS
 - LEGACY SYSTEMS
 - NEDIS enterprise ground concept

- Cloud tech: public cloud- disposable copy, running services for access and discovery and the machines of NOAA do not have to support spikes in usage. There can be a security boundary and have a private government cloud for secure users
- Assessment
 - Current state of data management
 - Aerospace study
 - Observing system of record DM survey
 - Progress measurements
 - EDMC reporting
 - DM DASHBOARD: prototype
 - Metadata sources across NOAA
 - Inventory
 - Quality of metadata
 - Provide feedback to data providers
 - Type of service offered
 - How many records point to an online service
 - What standard used?
 - Feedback from users
- Data lifecycle
 - Planning and production activities
 - Requirement def.
 - Planning
 - Development
 - Deployment
 - operations
 - Data management activities
 - Collection
 - Processing
 - Quality control
 - Documentation
 - 'cataloging
 - Dissemination
 - Preservation
 - Stewardship
 - Usage tracking
 - Final disposition
 - Usage activities
 - Discovery
 - Reception
 - Understanding
 - Analysis
 - Product generation
 - User feedback

- Citation
 - Tagging
 - Gap assessment
- Responding to the SAB (science advisory board), draft in progress
- There is a concept of operations: set up like a concept map
 1. Requirement guide for the data producer
 2. Then they follow a DM plan
 3. Then they generate the data
 4. While creating metadata
 5. Preservation through archive with an OAIS reference model
 6. Data access service will then be connected to a
 7. Catalog service thru
 8. A dashboard
 9. And NOAA leadership should be able to access dashboard
 10. From the data access service the data user can access through tool
 11. The result will then be some sort of product from user which will then be linked to
 12. Dashboard.
- DOI available? Fiscal 2013. Make sure that the landing page that links to DOI is managed in perpetuity.
- Question on catalog service? ESRI GEOportal server: open source and free. There is a GEOportal working group.
- Question about rapid movement with security and tech evolution: how do you reconcile break the boundaries to get security in line with web publishing today: federal govt. wide effort and in terms of software, have some pre-blessed from CIO's and offer and secure it through the server and that will help with access. Is there permission for the cloud? There is a shared posting project (if you don't want to use server) there is prototype in the works.
- **Question: dataONE: some interaction, more with EarthCube. This vision is in line with NSF guidelines----- who asked that question. Meet.**
- Ongoing stewardship: documentation can happen at several points in lifecycle. As much as possible automate-(metadata scanner?) suggesting archive work on accession.

New Speaker:

- Documenting NOAA data
- Using ISO 19115 for documentation
- 19139 xml representation standard
- Metadata means different things for many departments, need to build a common vocab.
- Documentation important. Brought up Climategate: failed to be open enough about their work.
- Well documented data sets lends credibility
- Users have been changing
- Data preservation is communicating with the future- the data needs to communicate independently and be understandable
- Using rubric to establish a baseline
- Promote and highlight good example
- Support training!!! For improving documentation

- Teams for documentation- especially station history
- Participation and support in various organizations- these help the line offices
- Line office
 - Identify expertise: data stewardship teams
 - Data collector/data provider
 - Standard expert
 - Steward
 - Users
 - Assess current documentation
 - Discovery, use, understanding, initial evaluation of collections datasets, and services by standard experts
 - Legacy, discover, metadata- add in high quality metadata
 - Diagram helps to understand use and develop strategy
 - See how complete metadata record is
 - Create and improve
 - Data providers: spirals, rubrics, wiki
 - Standard experts: consultation/guidance
 - Users: use cases/needs
 - Publish and preserve
 - Stewards
 - Users
- Metric: percentage of records of scores of 25% completeness
- Standard evolution must be applied to legacy documentation
- Multiple dialects within the multiple communities: the documentation language is sort of the same, and the stds. Often overlap but there are many: ISO, NETCDF, THREDDS, KML, WCS, SENSORML, DIF, etc. impt. To understand relationship between and translate.
- How do we improve?
 1. Initial content
 2. Scientific questions: new requirements, new use cases
 3. Check back with users
 4. Check back with data collectors/providers
 5. Standard guidance/implementation
 6. Metadata content: independent of standard
- Potential spirals
 - Discovery
 - Identification
 - Connection
 - Extent
 - Distribution
 - Understanding
 - Acquisition Information
 - Text searches
 - Content information
 - Quality/lineage
- If you don't understand instruments or quality of data, spiral is malleable.

- You can use a rubric to understand progress within each spiral or goal, and the completeness level will allow you to understand the documentation is improving and a way to measure improvement. Rubrics are changeable.
- Metric calculation system
- Current test cases in the calculation system
- Leadership model: positive deviance: says that if you want to create change, you must scale it down to the lowest level of granularity and look for people within the social system who are already manifesting the desired future state. Take only the arrows (people) that are already pointing the way you want to go, ignore the others. Id and differentiate those people who are headed in right direction. Give them visibility and resources. Bring them together. Aggregate them.
- question: who owns the metadata and who provides the stewardship for, then does NCDC archive and share? Archive struggle. Granule metadata, these will be in the file rather than listed in the catalog. How does the granule metadata fit in with the rest? Want an ongoing system to archive, use DOI's to link. STORE EVERYTHING! The question is about whether to use a DOI at the granule level- use a DOI higher and use UUID at granule level.
- Question about oracle: error number: google! Google will be the system to resolve DOI problems in the future. If there is only one hit, then maybe we should not be preserving the data set. BUT MAYBE THEY DON'T KNOW ABOUT THE DATASET BECAUSE NOT DISCOVERABLE.

New Speaker:

- NCDC metadata activities
- Pillars of excellence
 - Station history
 - Collection
 - FGDC
 - ISO
 - DIF
 - Archive
 - Fixity: checksums (not quite a dig. Signature)
 - Inventories
 - Granule
 - netCDF: CF and Attribute conventions for dataset discovery (ACDD)
 - Discovery
 - CLASS Descriptive Information
- Accomplishments
 - NCDC metadata working group
 - Collection level metadata dialect
 - Data governance policy/procedure
 - Data & metadata system inventory
 - Terminology and commonalities
 - Metadata automation
- Currently in progress
 - Metadata versioning policy/procedure
 - Regional partners coordination

- Interagency/international collaboration

New Speaker

- NODC ocean data archives and metadata
- Scope: many different types of data to archive
 - Deep ocean to inland coastal waters, great lakes
 - Physical, chemical, biological observation
 - Space based ocean observation and products
 - Ocean model outputs, ecosystems analyses
 - Photo collections, video collections
- Format agnostic data stewardship: many different formats are managed and the structure is flex enough to accommodate a digital file
- SIP (submission info package) ingest and AIP() metadata
 - Will move into web-accessible folder
 - Some ISO or FDGC metadata comes with file, often not in catalog but can be populated into web-accessible folder
 - Each package is given a Unique Identifier
 - Accession tracking database keeps up with the newer versions and improvements
 - Every accession, which will become an archival info package, has full suite of info in tracking database
 - Work in progress- xml blob element that stores additional metadata (ISO or FDGC) that does not have an ATDB element.
 - Every package structure is built- same structure for all, accession ID file system has two aspects: “about” and “whatever given from provider” and “translation of data, so this is a data product”- this is often what a user is more interested in.
 - All backed up, checksums (fixity), archival holdings are version controlled
 - Will new DOI's be given to new versions? Keep same id.
- DIP discovery and access
 - Discovery tool is same interface as the tracking database and ability to search is almost same as internal access
 - Web accessible folder indexed by google and it is the search info source through the geoportal server as well as data.gov.
 - Representations of the data- are data access points through the metadata through the webservice
- CLASS coordination
 - Working on incorporating this archive and its users
 - The data will also still come in for the in house archive and web accessible folder- offer greater and more robust metadata
- Diverse holdings
- Automated Ingest: high volume, low diversity
 - Negotiate with data provider
 - Thorough data appraisal process
 - 17,000 packages through acquisition tasks
 - Most have robust metadata that accompany
 - Requires upfront info to id all attributes they want- program like Bob wrote

- Non-automated ingest, low volume, high diversity
 - One time only data, non-repeating
 - Recognized as needing to be archived, require lots of scrutiny
 - Ingest takes time: hours to weeks or months
 - About 1/day.
 - Give guidance to data provider in order to provide proper metadata
 - Online submission system development process
- File level
 - Encourage data providers to use netCDF
- Legacy files present a great problem for implementing new metadata- so many files and so few resources. But if someone finds something that is not up to par, bring up to par while looking at it. Not well coordinated, but happening in spurts.
- Value added information- derived product? Keep them together but in a separate compartment of same package. Make sure have original data- this could be a source of criticism in response to customer concern.

New Speaker:

- Tools for creating and editing ISO metadata
- MERMAid
 - If used to FDGC- this is a good interim step.
 - You can export, see how metadata translates into ISO
 - This tool will go away
- CatMDEdit
 - Helps you to start thinking in ISO
 - Hierarchical structure visible and how things nest
 - Can load many file formats, and will populate some of metadata fields
 - Free
 - GUI tool
 - Nice html view
 - “baby iso tool”- no dash 2 ISO support. No search. No common catalog. No transforms.
 - Does not support xml attributes (ex: unknown date)
 - Nice interface to learn about ISO
- GeoNetwork
 - Robust tool
 - You can work by package or you can work on entire collection or go to xml view
 - Nice transition
 - Hierarchy and nesting
 - Supports 19115-2 ISO.
 - Profile support
 - Supports future catalog ISO
 - Platform independent (i.e. apple or pc)
 - Multilingual
 - Great validation??? What does this mean
 - Workflow set up
 - Ingest and export- batch

- Search interface
 - Free
 - Transform engine doesn't support transforms NOAA is putting out, but can go in between dialects
 - Can fiddle and will support XXL2- but customizable
 - Run on a dedicated system but if want multiple folks working on, run on server
- ISOMorph
 - GUI
 - Gives drop down lists and gives you direction
 - Supports xlink
 - Workbook allows for definitions of what goes in
 - Will produce xml at end
 - Can ingest xml also
 - Free
 - Pretty good validation, but not external afterwards
 - Platform independent
 - Nice dropdowns, but once generated metadata, must save metadata elsewhere
- Geoportal
 - Nice tabs
 - Auto-populate some fields
 - Asks for fields in human readable
 - Free
 - Integrated workflow for review or publishing metadata
 - Doesn't fully support stds, some missing, but constantly updating
 - Don't have vertical extent support if not widely used in GIS
 - Not biological support
 - No transforms for NOAA like xpath
 - Some x links, not all
 - Some xml attributes
 - Cannot edit auto populated fields unless go in to xml
 - Does not tell you what it removes- so can be dangerous for metadata
 - Nice search interface- what built for
 - Sourceforge. Com
- Altova Authentic
 - Plugin for web browser or desktop instance
 - Apply xml template and apply skin over, highly customized GUI
 - Built in validation- will tell you in red text why wrong
 - Free
 - Supports any standard
 - Have to develop skins externally but must export out to save
- XMLSpy
 - If deal with metadata a lot, these are types of tools you want
 - Must purchase
 - Apply skins- for people uncomfortable with xml
 - Any standard
 - Batch edits

- Versioning control- project synch
 - Graphics from schemas
 - Performs transformations
 - Validation
 - Tie into other tools
 - Everything in one tool- you get it all, but does cost.
 - Altova.com
 - Once given one training – easy to use. Worth it.
- Oxygen
 - Nice navigation theme
 - Lets you know hierarchy in ISO
 - Autocompletion
 - CSS, so can make a skin
 - Support any standard if you have schema, using multiple stds at one time
 - Validation
 - Transformations
 - Easy to edit
 - Batch functions
 - Bummer: must set up validation and transformation scenarios
 - Costs: comparable to spy.
- ISO Metadata editor review
 - Good place to look into different tools
- Other tools are transforms
 - A new docu created based on content, but does not guarantee validation- must edit after
 - Crosswalks one standard to another
 - Automation is way to go
- Docucomp
 - Record services and _____ services
 - Check validation and apply additional validation
 - Transforms for rubrics
 - Transform one dialect to another
 - You can baby step your way into ISO
- EPSG registry
 - In metadata you can point to a vetted registry to reduce redundancy

New Speaker:

- Mercury: distributed Metadata management, Data Discovery, and Access System: an internet tool to access
- NASA'S ORNL DAAC: distributed active archive center for biogeochemical and ecological data and models useful in environmental research
- 4 areas of NASA's terrestrial ecology program: field campaigns, validation, regional and global, model archive
- ORNL DAAC and NASA's 12 Earth observing systems
- Multiple distribution mechanisms
 - FTP
 - Metadata search catalog-mercury (FGDC compliant)

- Nasa Earth Observation system-wide clearing house
- Spatial data access tool (WCS) WebGIS
- Thredds Data Server
- MODDS
- DIVERSE GOALS:
 - Assemble data bases
 - Education and outreach
 - Technology
- Data management challenges
 - Diverse science
 - Custom models
 - Remote sensing
 - Success in science based on publications.
 - V. little metadata in raw data files
 - Mercury is based on open sources: main search engine
 - Can search metadata (archived products) or advanced product searches
 - Use metadata editor then mercury builds an index
 - Then users query this index
 - Within the index and metadata are links to the data and data providers
 - Metadata repository. Data lives all over the place
 - Two models: virtual- organizes info all over internet; virtual aggregate: harvests info from existing formal DBMS
 - Provides a single portal to info contained in different data management systems
 - Provides fielded, spatial and temporal search capabilities
 - Provides custom interfaces
 - You can search data sources- ex: FLUXNET
 - All collection level metadata but the granule metadata is included but further down.
 - New internal and external metadata editor in development- covered by Viv. Hutchinson.
 - This is now developed as a skin
 - A specialist only has to go through and get the minimal amt. of info.
 - Minimal pressure on data provider
 - GCMD based keywords for metadata searching
 - Back end indexer?
 - The DAAC assigns a DOI and then register with a handle on the front.
 - DataOne = ONE Mercury! Same basic as ORNL. Up and running!

New Speaker: Viv Hutchinson: USGS Core Science Metadata Clearinghouse: an Overview

- Discussing another use of mercury
- New data lifecycle

- Plan
 - Acquire
 - Process
 - Describe (metadata, documentation)
 - Analyze
 - Preserve
 - Publish/share: clearinghouse
- Data documentation workflow: making documentation easier: dashboard
- Clearinghouse helps to get data out into the web, data.gov and dataONE
- Clearinghouse
 - An open access, online metadata repository
 - Can locate, evaluate, and access earth science data
 - Provides a central distribution point for metadata record sharing in USGS
 - Participates in mercury consortium through ORNL
 - 103+ thousand records- mostly biology, ecology, geology, geospatial, + descriptions of tools, software, etc.
 - Standards
 - FGDC (federal geographic data committee)- conversion via crosswalk
 - Ecological metadata language EML
 - Dublin core
 - GCMD DIF
 - Other standards can be used as well
 - Users and partners- varied and large
 - How partners participate: have in xml format in web accessible folder and set up a harvest, generally about 1/ week. Every week are reharvested- handles updates, or made changes.
 - Advantages that the partners have to hold on to their metadata
 - Quality control by USGS FORT: passed through metadata parser to check record validity
 - Clearinghouse search interface- v. mercury looking
 - Can email search to yourself
 - Can save as RSS feed or subscribe- can rerun later
 - Can view full metadata or can get data- takes you to the center or landing page to get to data set
 - Get abbreviated metadata record or can view record in either FGDC or EML
- Metadata tool with oakridge- uses everyday language to ask questions about dataset and ease overwhelming process for scientists- when you get to end of form you can preview your record, save to your machine, use a specific standard, or save and submit and this will go to quality control people who will then put in clearinghouse after.
- Idea is that this is expandable and will be able to kick out an ISO record depending on the question
- Visualizations: this will show you the content graphically

- The relationships to other efforts: data.gov; global forest info service thru RSS feed, dataONE is a member node. NSF initiative. Global cyber infrastructure. Distributed framework- ability to search all data in one place. SO RAD.
- dataONE has various tools in system so that there is a workflow- various tasks can be done with
- question: several instances of mercury- is each one for a separate collection- just using same basic infrastructure to harvest metadata- different from instance to instance.

New Speaker:

- Ken Roberts: The advanced tracking and resource tool (ATRAC) 2.4 and metadata creation tool
- ATRAC
 - Web based tool for managing and tracking data archiving projects at a data center
 - Provides a central interface to enter project info and manages it for project status display and
 - Info flow: built in controls that control the flow of info. This info can be used to create standard documents such as XML and PDF
 - THIS is not a docu management system, just used to create
 - Search page: 1. Timeline view 2. Grid view
 - Metadata form: usable without knowledge of stds.
 - Validated input, form is intended to be part of the metadata process
 - Included elements: base format
 - NCDC REQUIRED SET
 - Additional
 - Browse graphic
 - Additional keywords (place, stratum, platform, instrument, project, resolution)
 - Cross ref
 - Applied quality measures
 - Processing description
 - Processing documentation
 - Input dataset citation
 - Controlled vocab. S
 - Metadata import
 - Import from other forms in ATRAC: archive form or submission agreement
 - External xml files, transform via xslt, store in ISO in archive
 - Metadata export
 - This tracking system is hype! Color coded legend, timeline, ingest status, etc. also the grid view allows you to organize your search based on the columns.
 - Validates the record- metadata record is even v. automated in the areas that you will complete

- If you run a rubrics test- it will actually catch mistakes and show that the lineage is not included. This allowed ted from boulder to catch a mistake of NCDC
- New steps for future
 - Updated submission agreement form
 - Integration with NCDC ingest configuration (SIPGenSys)
 - Management of shared components with docucomp
 - Improved XSLT's
 - Metadata file import
 - Service metadata (19119)

New Speaker:

Linda wayne: lwayne@fgdc.gov

- **FGDC** Metadata Summit & FGDC Metadata support strategy
- THE summit was mostly about the ISO training: purpose to bring everyone to table and see where people were with standard opportunities with standard and issues the std was causing
- Breakout groups:
 - Implementation
 - Policy
 - Directive
 - Timeline
 - Guidance
 - Website
 - documents
 - Workflow solutions
 - Some agencies said, until they have to do it, they won't. but some created agencies
 - Make it part of regular process
 - Training for specific types of roles
 - Curriculum
 - Materials
 - events
 - Offer training through webinars
 - Software specific
 - Ideal: online help desk and community workspace at FGDC
 - Need to provide current stds and processes
 - Tools
 - Implement so work can get done
 - Metadata editor review
 - ISO explorer
 - Engage with vendors to enrich tools and software
 - Work with ISO, to help develop
 - Develop portals that can publish ISO metadata
 - Communications
 - Get active with metadata working group

- Everything needs to be in one place- website, in order to consolidate and provide guidance and tools that are up to date
- NOAA has been the lead in ISO training
- What does the group think?
 - What does FGDC staff need to do to move ISO forward?
 - How to improve ISO once you get there
 - What kind of tool can evaluate the translation from FGDC to ISO or any other standard?
 - We must build on the wiki- it should be more comprehensive.

New Speaker:

- NASA's global change master directory- works for wiley info systems (GCM)
- New interfaces: KMS, MWS, and NextGen
 - KMS = keyword management service. Xml.
 - RESTful webservice that provides dynamic access to all GCM keyword sets in
 - SKOS
 - RDF
 - OWL
 - CSV
 - REST paths:
 - Concept resource
 - Concept-root
 - Concept- pattern
 - Concept-?
 - MWS= metadata webservice: direct access to DIF's and web services, now searchable. Xml
 - NextGen= preview- new.
 - New design and structure
 - Improved access to data, services, and ancillary descriptions
 - Advanced "tree" keyword search refinement via [google web toolkit](#)
 - Integration with MWS/KMS WEB SERVICES IN BACK END AND DOCbuilder
 - <http://gmcd.nasa.gov/nextgen>
 - Programmed in JAVA with [spring framework](#)
- Demo
- Usage Metrics
- Feedbacks: need a DIF to ISO translation.

New Speaker:

- CEOS WGISS integrated catalog: A catalog for earth observation satellite data
- GEO: group on earth observations
 - Initiative started 10 years ago, ministerial level- international member
 - Make more available and apply data to a range of applications
 - Remove some barriers to using and accessing international EOData

- GEOS: global earth observation system of systems wanted a common infrastructure- GEOS registry populated with the wide range of EO resources
- CEOS
 - Committee on earth observation satellites established in 1984
- WGISS
 - Working group on information systems and services objective statement
 - Promote collaboration in development of systems and services based on international standards that manage and supply earth observation
- Why CWIC?
 - Huge number of data resources
 - Data systems that support the satellite data systems
 - Long history of trying to improve accessibility of satellite data resources
 - Allows the agencies to work together and move forward together
- CWIC implementation
 - Uses a mediator-wrapper architecture to access the heterogeneous systems of the CEOS agencies
 - Uses the GEO catalog standard to interface with clients
 - Search responses returned using ISO 19115 metadata
 - Approach addresses the WGISS directory/inventory (collection metadata model)
 - Directory/inventory search: each CWIC collection has a unique DIF with a unique DIF entry ID
 - User can select science keyword, platform and instrument keywords as well as a spatial area and time range to narrow down to a smaller list of collections of interest
 - Currently, both the data center and the data collection name must be specified in the inventory search
 - Metadata: ISO 19115: limited subset when searching
 - Dataset name
 - Data center
 - Spatial
 - Temporal value
 - Full set can be returned in original data format

New Speaker:

- Curt Tilmes: US global change research program national climate assessment global change information system
- 1990 global change research act: wanted to coordinate many agencies
- Global change vs. climate change. Global change looks at climate change and how it impacts the global community
- To meet fed requirements: utility, integrity, objectivity
- Transparency, authenticity, etc.
- Track chain of custody of data: provenance
- Need to have a website that offers an interoperable data system but not going to host data
- One of the goals is leading people to data, but not hosting.
- Want to organize information about global change

- Interagency information integration
- This is a way to prove and provide a metrics for how useful the data is

So this is mainly a provenance website for data and agencies and publications?

Is this not the same thing as DOI'S?

I may be naïve. But it seems like many folks are talking at the moment about have a centralized, organized source for data. However maybe what he is saying is that the site is really just about making sure people understand how these agencies are linked. And they need to be identified, the agencies.

NCDC metadata workshop
8/15/2012

Speaker:

- Speaking on the changes with CLASS: the archive for all of the NOAA centers
- About the architecture of interfaces
- Must have enough metadata to have repositories and metadata archives can pull out of this main archive
- What does CLASS need vs. what does the data center need?
- Have object ID's.
- Curious about how much metadata they must include
- Won't have any metrics to learn about your data if you do not include metadata and DOI's.
- They need who, what, when, where. In order for data center and others to access data.
- Long term solution: abstract archive, interfaces, outside archive is where all of the action occurs- this allows the data consumers to not care where the data lives as long as they can locate it. The abstract archive can use different storage units but all happens outside context of interface--- this is a service model
- Conops:
 - Satellite provider
 - CLASS extracts DI and metadata (robust but also redundant)
 - CLASS provides access via CLASS web
 - Metadata not exposed
 - Data center provider
 - Etc. slide went fast
- CLASS wants unique ID, who provided, location, time, data group, creator id
- Struggling to provide the balanced amount of metadata so that data can be found but also not take on the role of stewardship that a data center will take on.
- CLASS doesn't want to be in metadata business or website business, they want to maintain being an archive
- Machine to machine searching- only will search metadata provided by data center vs. satellite data tells CLASS tells what to pull from the data

- Data centers with their own access centers/portals, their own archives- goal is to have data centers getting back in to the catalog/website business and have only one archive so that the data centers can get out of the archive business and have only one archive
- Common storage service: cloud access: taking often used or newly arrived data and putting it on a spinning disk to reduce effort on the data center- many can read and access and download quickly- this is referred to as elastic computing services?

New Speaker:

- ECHO system and data model overview
- This is a NASA middleware layer between earth science data and users via service-oriented architecture. Designed to improve the discovery and access of NASA data.
- Fully adaptable, ready for change, and looking to support ISO 19115 and various standards
- Webclient: REVERB: primary web based client for discovering and ordering cross-discipline data from all of echo's metadata holdings
- ECHO service registry: a bit buggy, but allows users to id collections or granules and export to the user, only with ECS style data centers.
- ECHO has about 14 providers- who then have collection and granule level data
- ECHO catalog is an oracle database
- Data flow
 - Rest API for searching
 - When doing ordering, have to download and submit form- no documentation for this part of the data retrieval service
 - Also an ESIP connector to try this open search
- Usage metrics
 - Try to exclude spiders
 - Chance for all data partners to express problems and what trends are
 - This is REVERB granule queries
 - Struggle with query performance- most queries answered in under ten seconds, skewed based on types of access- some people want all data at a moment, which slows process
 - Astrid volcano team started using API for search ordering- shot through the roof on query numbers
- Question: what are plans for scaling? They are at capacity. They are having to ask users if there are bugs due to huge data downloading
- Looking to:
 - Format translators
 - Format indexers
 - Retaining pristine format
 - Validation
- There is a challenge in validating data and qualitative analysis- not their job she said but also thinks they should practice this.

New speaker:

- Ted Habbermann, NOAA/NEDIS/NGDC(from boulder): friends with mary marleno
- ECHO and ISO are dialects and he is interested in mapping between. ECHO has some info that is outside ISO, most related to instruments and processing

- Suggestion to NOAA would be to support in ISO stds. Dash 2 is behind in revisions- mostly dealing with acquisition revisions, still can participate
- Impt for NOAA to participate in std development (between 19115 vs. DASH 1- lots that NASA and NOAA needed that weren't included)- quote 'do stds before they are done to you'
- Docucomp:
 - Database normalization is the process of organization the fields and tables of relational database to minimize redundancy and dependency
 - OPeNDAP: getting stuff out of files and into files. There is a model that id's many types of data objects- persistence vs. transport
 - Roles, identifiers, and references- this is ISO xml. The roles have references to the objects and comes from UML. Difference between ISO xml representation and every other xml representation- the reason is so that you can easily reuse content.
 - Most folks deal with collections of metadata records- collected over time.
 - Database plus a **restful service** aka: component database
 - Component resolution: the record set is the collection of related records, an unprocessed directory- the reference is an xlink
 - When record resolved, goes into a directory called ISO, xlink resolved.
 - Before users get to this, they should only see resolved items, however users don't necessarily care about contact and resolved info- smaller because ISO citations take up a lot of space. So if using 6 or 7 keyword lists from GCMD that adds a lot of bulk for the user
 - Getting there:
 - Existing WAF (URL) →
 - Harvest ->
 - NGDC processing directory →
 - TRANSLATE →
 - Iso_u
 - RESOLVE →
 - With component database →
 - ISO
 - Iterations:
 - TRANSLATION PROOFING: id un translated content fix
 - RUBRICS: evaluate completeness and improvement opportunities
 - VALIDATION: schematron and schema validation
 - LINK CHECKING: check all online resources
 - MULTIPLE VIEWS: translation of ISO to multiple views
 - CONSISTENCY CHECKER: find repeated content and candidate components: table of contents of records for metadata

- manager, in order to improve metadata
 - DOCUCOMP: component library
 - METADATA EDITATOR
- The wiki page will be linked to translations to act as a resource and guide for information and can be changed . integration of rubric into wiki.
- Bottom of table of contents are generated in processing system, this will let you know which URL's are bad or bad xlinks. Info provided to metadata creator, this will then be put in dashboard so time history of these things will show up to help monitor
- Supporting legacy portals = FGDC = an xml.
- Discovery FGDC record links back to the ISO, might be in metadata section? url back to it.
- UUID report shows all components used in record set, in bottom shows all of the components and when they are hit- these are ISO xml. Not validated until reinserted
- Unresolved records are dynamic- once resolved those are the ones you want to archive, not the unresolved- great dynamically driven tool for maintenance

New Speaker:

- How do we start implementing ISO?
 - Training
 - Automation- know what tools are available
 - Transforms- located in many different places. Baselines- bio to ISO, remote sensing to ISO, etc.
 - Cross walks define how transforms work
 - In xpat2- need a processor in architecture, built in to some programs like oxygen and spy.
 - Should start looking at what is available already for automation- before we get into subgroups into developing on our own- because autonomy ends in having to change- wiki born out of needing to do a brain dump so that mistakes can be communicated ---- but communication should be broader in order for all agencies to know
 - Phases of implementation
 - NODC transitioning to ISO
 - OAR already implemented ISO and starting to write best practices- start metadata creation and now have a workflow set up
 - Ted far along, how can we make ISO better? Past implementation.
 - Mixed bag of where everyone is?
 - Web accessible format- in ISO. If harvesting from an FGDC or DIF, can also look at the record in ISO- but rubrics will only be in ISO.

- CLASS is harvesting ISO from 3 data centers and is integrating into display pages – collection level metadata, but granule metadata is inaccessible because within file- should be exposed within machine to machine new program- heap of work- the multiple formats will make the challenge all the much harder.
- Do I have vocabularies?
- When do the migrations start?
- How to attack legacy
- ISO is a complete shift- lots of power and capability to leverage, how can we exploit?
- If we use xlinks- what is the best process for managing these? These are management practices. What systems? What am I translating to? Do the transforms already exist? Sourceforge.com.
- Updating content so make more user friendly and then mapping into whatever xml structure is useful for management
- Changing workflow to use new tools- thus inserting more steps into workflow, also training
- Should you edit then translate or vice versa?
- Edits for translation vs. content improvement. Different if you can translate into ISO without loss, then go ahead and translate and worry about improving content later- Ted is suggesting this is a delay tactic. The challenge? Improve content in a system you know or shall you wait to improve content until you get into a system that is less familiar like ISO? Perhaps more political/cultural changes rather than IT. Suggestion of a workflow decision tree, useful mechanism for people who need to move forward.
- Controlled vocab is key in order to facilitate search- decision made in implementation phase.
- Point made: rather than being so concerned about controlled vocab, go ahead and translate (more automatic) and allow the controlled vocab development to be a task put off until later for improvement.
- My question: so if they are suggesting that you use a new standard rather than improving your quality, what's to say that there won't be a new version of ISO that needs to be implemented before the content is amended. However, they say the translation should be easy into another version. Another point with this is translation loss.
- Answer: there are now approaches to what may need to be added or perhaps what is lost that isn't needed. What they are worried about is the unknown untranslated items.
- Difference between extension vs. profile (this is in response to the NASA flavored ISO conversation). Jackie says this is where schematrons come in. when you check the

validation- you can add additional constraints to check on the schema accuracy.

- Talk of organization skins- you can edit your transform to edit so that if your organization standardly misused the standard, you can custom
- Thinks that code lists should be shared online, so that the development does not become redundant among agencies
- Data.gov plays an imp. role because CIO's pay attention to it. Discussing inconsistency in agency names, same as Chris Borgman talking about how each person should have their own ID that links all of these different entities that are actually one agency. Imp. to improve content of metadata. GCMC has an approach to handle this issue. This is where ideas like consistency- encourage creators to do consistency, rather than merely creating a thesaurus of all of the agency.
- A record and a record type are how you custom ISO- you can add specific attributes, a possibility is to add ECHO xml in.
- Wiki tells you about the new capabilities that will show you what you might have to change or what the new services are.
- If human reads, put in distribution, If a computer reads put in services
- aggregation info: cross references and citation: also cite up to the larger collections (parent vs. child). Where site project and program.
- Identifiers- reoccurring theme. Not going to be solved, but will be evolving.

Interesting point towards keyword search: if you have a multilingual audience, make sure that keywords are available because non-native English speakers recognize this more than free text. However ISO is supposed to resolve this issue by being an international standard.

NEED TO LEARN DRUPAL

New Speaker

- Metadata usage: global change master directory
 - Used to discover datasets: uses directory interchange format
 - Data held in database and put in an xslt
 - Adaptation of the DIF is SERV
 - Ancillary descriptions are about campaigns, platforms, etc.
 - Another use of the directory is for climate diagnostics
 - Different from a dataset, describes a diagnostic or visualization someone created

- Doc builder is the tool for describing a dataset, trade ancillary descriptions, or describe a SERV.
 - You can customize
 - User account required
 - If a record in production, public, or cache system(not publically available), cache is auto saved for 90 days
 - Doc builder acts as a metadata editor?
 - You can edit xml if want, but most just edit fields
 - Scientists like because they don't need special training
 - Types of metadata
 - Discovery
 - Granule
 - Collection
 - Archival
 - All work for interoperability

DISCUSSION

- "Institutionalize metadata before it institutionalizes you"
- What is the minimum fields required for sharing?
- Harvesting techniques will be changing, so they can be documented
- Csw were not std and stable enough so geoportal has a much better reputation- this breaks us out of clearinghouse mode? Only people without csw's will have to deal with all of the harvesting
- Need to go to CF to get a standard for documentation to move forward- Ted thinks will be difficult because netcdf 3 and netcdf classic don't support groups in the attributes
- Unidata people are fine, libcf people (tools: IDV, FARROT: netcdf tools that assume no groups: code development problem) – only get two
- At the granule level, how do that automated systems search? Recognizing whatever vocab presented, there will be something unique from each center.
- ISO 19117 is the new standard that is higher quality.

New Speaker:

- Ted habbermann: metadata tools
 - Interested in how organizations work and evolve
 - Finding tools that work are really important- if we give a task we should give a tool that works
 - In early days of FGDC there were no real xml tools
 - Why an xml management tool? Connects to a schema (gives you the structure of the xml you are working with, required attributes, etc)
 - This allows you to use content completion
 - Schema tells tool what can go here
 - Schematron: business rules validator- makes sure you have a keyword from the GCMD, can test code list formats, can make sure that values fall within the correct range
 - XXL are completely built in
 - CSS built in also
 - Makes easy to edit

- Schema can be annotated, so can have access from the editor
- Question: who should be creating metadata- not the scientist!
- When talking about data stewardship teams (back in the day, there were metadata librarians)
- There needs to be a process that is assisted by a tool. The tools help the interaction between the scientist and metadata editor.
- HEATED DEBATES!
- CSS is a layer on top of a commercial tool. Could you put ATRAC on top of oxygen as a skin? How long would this take? The nice thing about the implementation is that you can look back and forth between the two stds. With a mere click of the tab.
- Xml is a way to represent the information. We can get lost in the weeds of particular representation, when we look at document management, we want tools that abstract away from having to know code. You want a what you see, what you get edit.
- Response to this has to do with cost: Ted is arguing for a curator, others are arguing for a tool that a scientist can use. Ted is saying that a human is the best tool.
- Just likened a data curator to a secretary!
- Ted replied: in metadata land, scientists don't want to do this task.
- They want everything automated. Get rid of humans!
- The impt. thing of the xml tool is that it supports all of the dialects such as ISO, FGDC, thredds, dif, ETC.

NCDC Metadata Workshop
8/16/12

Discussion:

ESIP: federation of earth... providers: brings govt, academia, and commercial to bring about opportunities to bring about a better education initiative on earth science.

- The Commons: formal publication: under ESIP, mechanism to assign DOI s to their internal papers (ex: one on data citations)
- ESIP works on data training- data curators? To prepare somebody to go into this line of work- what should they know how to do? Great opportunity for data centers to reach a different audience.
- GOSAR: granule + metadata, + provenance metadata included, so granules may be reprocessed by later scientists

- We need to be more efficient and know what is out there. How do we communicate what each organization and people within organization or group task- this would help the process move along- better communication: WIKI, mailing list
- Phil wonders, how do we involve the data providers: in order to get the metadata into the file, there needs to be some sort of incentive- for standardization of data management plans.
 - What are the incentives?
 - Awaiting approval
 - Publications
 - Find the people who want to do it, and make good examples out of them
 - Talking about how to implement: ted says from the bottom. He says the community needs to drive, rather than their top.
 - Choosing the battles and provide evidence that time must be allocated for this metadata work will be done
 - Impt to share the processes across the NOAA data centers and learn from one another
 - If you don't document the data as it is happening, then so much is lost.

Who are main data providers: centers or researchers?